

第40回

インシリコ・メガバンク研究会開催のお知らせ

2013.12.06 [金] 開場 16:45 開演 17:00~18:30

[会場] 東北メディカル・メガバンク機構2階会議室1

[演題] 一般化t統計量の漸近的性質の考察とその実データ解析への応用

[講師] **小森 理**
(統計数理研究所)

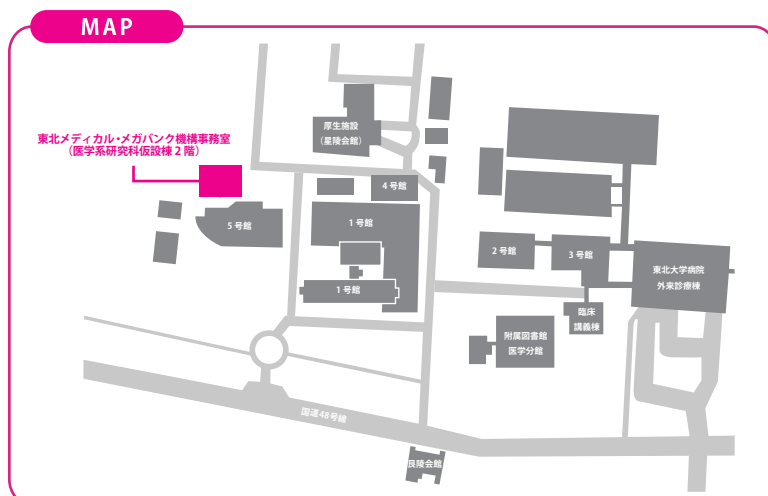
[問合せ] 東北大学 東北メディカル・メガバンク機構
ゲノム解析部門 長崎研究室事務
TEL 022-273-6051/ E-mail nagalab-jimu@csml.org

高次元データの解析において臨床医療またはさまざまな病気の診断に有効な変量（マーカー）の探索は近年ますます重要になってきている。遺伝子発現量のデータのような高次元の連続量のデータ解析においては、t 統計量やc 統計量 (AUC) などが変数絞り込みの段階でしばしば用いられる。今回我々はこの検定統計量の一つである t 統計量に着目し、単変量ではなく多変量の線形結合を考慮することでこの t 統計量を判別問題に応用することを考える。

t 統計量に対し生成関数であるU関数を考えることにより、さまざまなt統計量の一般化が考察できる。一つの結果として t 統計量、c 統計量 (AUC)、Fisher 線形判別、Kullback-Leibler divergence との密接な関係が明らかになった。またこの手法の実データへの応用例として一般化t統計量にL1ペナルティをつけた Lasso タイプの手法を提案する。

この手法の一つの特徴として、あらかじめ有用であることが分かっている変量（発症リスクに関連が深い変量）を固定したうえでその他の変数の選択ができる点があげられる。シミュレーションと実データ解析においてその有用性を検証する。また将来的にはSNPのような離散的なデータに対しても適用可能な統計的手法を考えていきたい。

世話人: 田宮 元 長崎 正朗



※本セミナーは医学系研究科系統講義
コース科目の授業としての振替可能
なセミナーです。